METHOD AND APPARATUS FOR REACHING AGREEMENT BETWEEN NODES IN A DISTRIBUTED SYSTEM

ABSTRACT

One embodiment of the present invention provides a system for selecting a node to host a primary server for a service from a plurality of nodes in a distributed computing system. The system operates by receiving an indication that a state of the distributed computing system has changed. In response to this indication, the system determines if there is already a node hosting the primary server for the service. If not, the system selects a node to host the primary server using the assumption that a given node from the plurality of nodes in the distributed computing system hosts the primary server. The system then communicates rank information between the given node and other nodes in the distributed computing system, wherein each node in the distributed computing system has a unique rank with respect to the other nodes in the distributed computing system. The system next compares the rank of the given node with the rank of the other nodes in the distributed computing system. If one of the other nodes has a higher rank than the given node, the system disqualifies the given node from hosting the primary server.